JOURNAL OF BACTERIOLOGY, Oct. 2006, p. 6858–6868 Vol. 188, No. 19
0021-9193/06/$08.00+0 doi:10.1128/JB.00421-06

Comparative Genomic Hybridization Analysis of *Enterococcus faecalis*: Identification of Genes Absent from Food Strains†

E. Lepage,[1] S. Brinster,[1] C. Caron,[2] Céline Ducroix-Crepy,[3] L. Rigottier-Gois,[1] G. Dunny,[4] C. Hennequet-Antier,[2] and P. Serror[1]*

Unité des Bactéries Lactiques et pathogènes Opportunistes,[1] Mathématiques, Informatique et Génomes,[2] and Centre de Ressources Biologiques,[3] INRA, Jouy-en-Josas, France; and Department of Microbiology, University of Minnesota Medical School, Minneapolis, Minnesota[4]

Received 27 March 2006/Accepted 6 July 2006

***Enterococcus faecalis*, a member of the natural microbiota of animal and human intestinal tracts, is also present as a natural contaminant in a variety of fermented foods. Over the last decade, *E. faecalis* has emerged as a major cause of nosocomial infections. We investigated the genetic diversity in 30 clinical and food isolates, including strains V583 and MMH594, in order to determine whether clinical and food isolates could be distinguished. Data were obtained using comparative genomic hybridization and specific PCR with a total of 202 probes of *E. faecalis*, selected using the available V583 genome sequence and part of the MMH594 pathogenicity island. The cognate genes encoded mainly exported proteins. Hybridization data were analyzed by a two-component mixture model that estimates the probability of any given gene to be either present or absent in the strains. A total of 78 genes were found to be variable, as they were absent in at least one isolate. Most of the variable genes were clustered in regions that, in the published V583 sequence, related to prophages or mobile genetic elements. The variable genes were distributed in three main groups: (i) genes equally distributed between clinical and dairy food isolates, (ii) genes absent from dairy food-related isolates, and (iii) genes present in MMH594 and V583 strains only. Further analysis of the distribution of the last gene group in 70 other isolates confirmed that six of the probed genes were always absent in dairy food-related isolates, whereas they were detected in clinical and/or commensal isolates. Two of them corresponded to prophages that were not detected in the cognate isolates, thus possibly extending the number of genes absent from dairy food isolates. Genes specifically detected in clinical isolates may prove valuable for the development of new risk assessment markers for food safety studies and for identification of new factors that may contribute to host colonization or infection.**

Enterococci are ubiquitous low-GC percentage gram-positive bacteria encountered in various environments, including animal and human intestinal tracts, soil, plants and water. They are found as members of the natural microbiota of a variety of fermented food products such as artisanal cheeses and fermented sausages and reportedly play an important role in food processing (23, 28). Enterococci have been consumed for centuries, and both *Enterococcus faecalis* and *Enterococcus faecium* species are subdominant members of the digestive microbiota in human. However, isolates of both species are emerging as major causes of nosocomial infections, including urinary tract and abdominal infections, bacteremia, and endocarditis in patients with severe underlying diseases or an impaired immune system (40). *E. faecalis* causes 60 to 80% of enterococcal infections. Less than 2% of infections are due to strains resistant to clinically relevant antibiotics, ampicillin and vancomycin, implying that other genetic factors are necessary for *E. faecalis* infection and virulence potential (30).

Phenotypic studies suggest that *E. faecalis* strains vary in their colonization and invasion abilities and, thus, likely vary in their virulence potential (14, 26, 39). Overall genetic diversity of *E. faecalis* has been reported using various molecular typing methods (for review, see reference 18; also J. C. Ogier and P. Serror, submitted for publication). While diversity might explain the presence of *E. faecalis* in various environmental niches and contribute to virulence, little is known about the identity and the distribution of the variable genes. Among the dozen of *E. faecalis* putative virulence factors reported (for a review, see references 27 and 35), sets of known and potential virulence factors (e.g., aggregation substance, enterococcal surface protein [Esp], cytolysin toxin [Cyl], and gelatinase [GelE]) are widespread among various collections of isolates, including food-associated isolates (3, 10, 14, 16, 17, 23, 33, 53). The findings that *E. faecalis* virulence genes are detected in food-associated isolates calls for safety assessment measures (10, 16, 23). Up to now, studies of *E. faecalis* genetic diversity have focused on just three chromosomal regions, known as the pathogenicity island (PAI) (43, 54), the *fsr* locus, and the capsular polysaccharide gene clusters (32, 41, 48). Recent sequence availability of the clinical *E. faecalis* V583 genome (46) and of the pathogenicity island of strain MMH594 (54) has opened the way to explore *E. faecalis* genome diversity using DNA array technology for transcriptome analysis (1) and comparative genomic hybridization (CGH). CGH between pathogenic and nonpathogenic or avirulent isolates within a single species has proven useful for delineating putative bacterial pathogen determinants (7, 8, 12, 13, 15, 37, 51, 56). Despite the

* Corresponding author. Mailing address: Unité des Bactéries Lactiques et pathogènes Opportunistes, INRA, Jouy-en-Josas, France. Phone: 33 1 34 65 21 66. Fax: 33 1 34 65 20 65. E-mail: pascale.serror@jouy.inra.fr.

† Supplemental material for this article may be found at http://jb.asm.org/.

TABLE 1. *Enterococcus faecalis* isolates by source of isolation and MLVA type

Strain no.	Other designation	Origin	Sample source	MLVA type	GelE activity	Reference or source
VE14002	V583	Clinical	Blood	17	+	50
VE14039	DS16C3	Clinical	Unknown	15	+	22
VE14040	FA2-2	Clinical	Unknown	1	−	9
VE14505^T	CNRZ137^T, NCDO581	Unknown[a]	Unknown	7	−	
VE14510	CHV1324	Food	Poultry	2	−	CHV (Versailles, France)
VE14512	CHV44	Clinical	Urine	11	−	CHV (Versailles, France)
VE14514	CHV212	Clinical	Urine	19	+	CHV (Versailles, France)
VE14518	CHV442	Clinical	Deep pus	11	−	CHV (Versailles, France)
VE14522	JJG1	Food	Cheese	11	−	INRA (Jouy-en-Josas, France)
VE14523	JJG2	Food	St. Paulin cheese	13	+	INRA (Jouy-en-Josas, France)
VE14524	JJG3	Food	Cheese	16	+	INRA (Jouy-en-Josas, France)
VE14531	JJG40A	Food	Livarot cheese	9	+	INRA (Jouy-en-Josas, France)
VE14532	JJGCIISA	Food	Livarot cheese	10	+	INRA (Jouy-en-Josas, France)
VE14534	JJGCIISC	Food	Livarot cheese	14	+	INRA (Jouy-en-Josas, France)
VE14535	CNRZ23G	Food	Rennet	8	−	INRA (Jouy-en-Josas, France)
VE14568	609	Clinical	Endocarditis	2	−	Health Sciences Center (Oklahoma City, Okla.)
VE14569	613	Clinical	Endocarditis	2	−	Health Sciences Center (Oklahoma City, Okla.)
VE14571	654	Commensal	Feces	3	−	Health Sciences Center (Oklahoma City, Okla.)
VE14583	CHV490	Clinical	Lung	16	−	CHV (Versailles, France)
VE14584	CHV597	Clinical	Blood culture	16	−	CHV (Versailles, France)
VE14585	CNRZ1388	Food	Egyptian cheese	4	−	INRA (Jouy-en-Josas, France)
VE14596	1313043	Clinical	Valve's body	15	+	25
VE14597	406022	Clinical	Aortic valve	15	+	25
VE14598	99196021	Clinical	Blood culture	12	+	25
VE14599	48273044	Clinical	Aortic valve	8	−	25
VE14600	MMH594	Clinical	Blood	18	+	34
VE14615	697	Food	Salers cheese	2	−	INRA (Aurillac, France)
VE14617	707	Food	Salers cheese	5	+	INRA (Aurillac, France)
VE14623	718	Food	Salers cheese	5	+	INRA (Aurillac, France)
VE14628	735	Food	Salers cheese	6	+	INRA (Aurillac, France)

[a] The origin of the type strain is contradictory in the literature.

widespread use of CGH, only few statistical approaches for data analysis have been developed. The most common data analysis methods use a constant ratio value as a threshold for assigning genes into either the absent or present categories. This threshold is usually empirically determined from a comparison of reference strains for which the gene distribution is known. This method supposes a high reproducibility between arrays. However, inherent differences between membranes and the fluctuating efficiency of DNA labeling are responsible for variation between arrays. The mixture model has proven to be a powerful statistical method to classify genes in a finite number of groups (11, 24, 38, 45).

To investigate genetic diversity of clinical and food isolates of *E. faecalis* on a larger scale, we explored 30 *E. faecalis* isolates from clinical and food origins for the presence of 202 genes. For this purpose, we performed comparative genomic hybridization on a focused macroarray containing 186 genes, mainly encoding exported proteins, and specific PCR on 16 genes. We applied a statistical approach based on the mixture model to ensure reliable DNA array data analysis that would distinguish the presence or absence of genes in a given isolate.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The *E. faecalis* strains used in this study are listed in Table 1. These include the type strain and isolates of clinical (15 isolates), food (13), and commensal (1 fecal isolate from a healthy volunteer) origins. Clinical strains were obtained from the Centre Hospitalier de Versailles (CHV), the Centre Hospitalier Universitaire of Lyon (25), or the Health Sciences Center at University of Oklahoma. Food isolates were provided by the CNRZ (National Centre for Zootechnical Research) collection at Jouy-en-Josas and the INRA collection at Aurillac. The commensal isolate studied was obtained from the Health Sciences Center at the University of Oklahoma. To assess the distribution of genes scored as absent in food isolates, 70 additional *E. faecalis* clinical, food, and commensal isolates from diverse locations (Argentina, Egypt, England, France, and the United States) were analyzed by specific PCR. A description of these strains is available upon request.

Diversity of the 30 isolates was first determined by multilocus variable-number tandem-repeat analysis (MLVA) using a modified procedure developed by Titze-de-Almeida and collaborators (57). For each PCR, 40 ng of total DNA was suspended in 20 μl containing 12 pmol of region-specific pairs of primers (MWG Biotech, Courtabeuf, France), a 0.2 mM concentration of each deoxynucleotide triphosphate, 1× T.Pol incubation buffer (10 mM Tris-HCl [pH 9.0], 50 mM KCl, 0.1% Triton X-100, 0.2 mg/ml bovine serum albumin) containing 1.5 mM $MgCl_2$ (Qbiogene, Illkirch, France) and 0.5 U of *Taq* DNA polymerase (Qbiogene). Reactions were performed in a Mastercycler gradient (Eppendorf). PCR was carried out as previously described (57) with 15 cycles of touchdown amplification-denaturation steps, except that 10 cycles of standard PCR with annealing at 55°C were added to increase the PCR product yields. PCR product sizes were analyzed on 1.5% agarose gels to determine the number of repeats. MLVA data were analyzed with the classification multivariate method of StatGraphics Plus, version 5.1. Equal weights were given to large and small numbers of differences in the number of repeats within a particular locus. MLVA analysis of the 30 isolates using *aceB*, *espC*, *efa3*, *efa5*, and *efa6* repeats (56) resulted in 19 distinct MLVA types (Table 1).

To investigate the distribution of genes enriched in clinical and food isolates,

a total of 70 isolates were analyzed by PCR in addition to the 30 isolates analyzed by DNA array hybridization. All strains were grown at 37°C without shaking in brain heart infusion broth. Detection of gelatinase activity was determined on Todd-Hewitt (Difco Laboratories) agar plates containing 3% gelatin and revealed as described previously (51).

Probe and primer design and macroarray construction. The macroarray was designed using 211 *E. faecalis* genes, of which 205 genes were from strain V583 and 6 were from strain MMH594. This selection includes three housekeeping genes used as positive controls (*gyrA*, *dnaN*, and *rpoB* corresponding to EF0002, EF0006, and EF3238 in V583, respectively). Three genes used as negative controls were from *E. coli* (*cheZ*) and *Saccharomyces cerevisiae* (YAL058C-A and YAL047C).

Oligonucleotide primers were designed to amplify PCR fragments ranging from 200 bp to 500 bp. Most of the primers were identical to those used in the V583 microarray designed by Aakra and collaborators (1). In order to perform two-stage PCR amplification (47), each primer was designed with a nonvariable adaptamer sequence at the 5′ end. The adaptamer sequences were 5′-TACCT TCTCGAGGGGAC and 5′-ACCCTCTCGTGGGCAG for forward and reverse primers, respectively. The first PCR step was performed in 50 μl of reaction mixture containing 20 ng of genomic DNA, a 200 μM concentration each deoxyribonucleotide, a 300 nM concentration of each primer (MWG Biotech), and 1.5 U of *Taq* DNA polymerase (Qbiogene). Amplification consisted of a denaturation step at 94°C of 5 min, followed by 2 cycles of 30 s at 94°C, 30 s at 56°C, and 1 min at 72°C and 25 cycles of 30 s at 94°C and 1 min 30 s at 72°C, with a final step of 1 min at 72°C. PCR products were purified after electrophoresis in 1.2% low-melting-temperature agarose using a DNA gel extraction kit (Millipore, Bedford, Mass.). Second-round PCR products were generated by reamplification of the purified first-round PCR products using the adaptamer sequences as primers. Reactions were carried out with 1 μl of 100-fold diluted PCR product, a 200 μM concentration of each deoxynucleotide triphosphate, a 300 nM concentration of each primer, and 2 U of Qbiogene *Taq* DNA polymerase. Amplification was achieved by denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 30 s, with an additional extension period at 72°C for 1 min. Before spotting, PCR products were precipitated with 2.5 volumes of absolute ethanol and resuspended in 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). PCR product concentration was adjusted to 200 ng/μl after DNA quantification by gel electrophoresis and Fluoroskan Ascent (Thermo Electron Corporation, Boston, Mass.). PCR samples were printed in duplicate on Nylon membranes (Hybond-N+; Amersham Biosciences, Buckinghamshire, England), using a Qbot macroarrayer (Genetix, Hampshire, United Kingdom) by the Centre de Ressources Biologiques GADIE (INRA, Jouy-en-Josas, France). After spotting, DNA was submitted to a denaturation step by treating membranes with 0.5 M NaOH–1.5 M NaCl, followed by a neutralization step with 1 M Tris-HCl–M NaCl, and membranes were rinsed four times with milliQ water. Membranes were dried for 1 h at room temperature and for 2 h at 80°C.

The specificity of DNA macroarray probes was analyzed by sequencing the 214 amplified DNA fragments with an ABI PRISM 3700 DNA analyzer from Applied Biosystems. In total, 186 *E. faecalis* genes, among which 180 were from strain V583 and 6 from strain MMH594, were considered for data macroarray analysis. Sixteen probes, which had to be excluded from the macroarray because of mixed PCR products (EF0511, EF0540, EF0605, EF0776, EF1420, EF2253, EF2525, PAIEF0047, and PAIEF0053) or putative cross-hybridization (EF0355, EF1896, EF1992, EF2250, EF2347, EF3256, and EFA0047), were analyzed using PCR. Their presence in the 30 isolates was examined by specific PCR amplification with primer pairs used for amplification of the cognate PCR probes. A full list of primers, PCR product sizes, and their nucleotide sequences are available as supplemental material (see Table S1 in the supplemental material).

DNA methods. Total DNA was extracted from 4 ml of late-exponential phase cultures as previously described (19). About 50 ng of genomic DNA was labeled with 50 μCi of α-^{32}P-labeled dATP (Amersham Biosciences, Orsay, France) using the RadPrime DNA labeling system (Invitrogen, Carlsbad, Calif.). Labeled genomic DNA was then purified from unincorporated nucleotides on Microspin S-200 HR or ProbeQuant G-50 micro columns (Amersham Biosciences) according to manufacturer's instructions.

Membranes were soaked in 2× SSC and prehybridized for 2 h at 42°C in hybridization buffer (6× SSC, 50% formamide, 1% sodium dodecyl sulfate [SDS], 10× Denhart's solution [0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin]) and 15 μg/ml of denatured and sonicated salmon sperm DNA. Labeled genomic DNA was denatured at 100°C for 5 min and hybridized for 16 h at 42°C. Membranes were washed using conditions of increasing stringency: 10 min at 42°C in 6× SSC–50% formamide–0.1% SDS, 10 min at 60°C in 2× SSC–0.1% SDS, 10 min at 60°C in 0.2× SSC–0.1% SDS, and 10 min at 65°C in 0.1× SSC–0.1% SDS. They were then sealed in Saran Wrap and exposed to Storage Phosphor GP screens (Amersham Biosciences) for 3 to 5 days. Screens were scanned at 100-μm resolution using a Storm Imager (Amersham Biosciences). Signal intensities were quantified using Imagene software, version 5.5 (BioDiscovery, El Segundo, CA).

Specific PCR amplifications were performed using conditions similar to those for probe preparation, except that the initial denaturation step was followed by five cycles of 30 s at 94°C, 30 s at 56°C, and 1 min at 72°C. This modification was introduced so as to allow amplification in the case of mismatches due to potential sequence variations between isolates.

To test whether the absence of the *EF1420* and *EF2144* genes was linked to the absence of complete prophages, we analyzed the junction regions of prophages 03 and 05 of V583 in strains that lacked the prophage genes. The primer pair OEF168 (5′-TGGGGTTAATCCATTTGACC-3′) and OEF169 (5′-TGACAGC TAAACAGTATGCG-3′) and the pair OEF170 (5′-AAATCTGTCATTCCAG CGAC-3′) and OEF171 (5′-TTTGACGATTACTCGTCGC-3′) were used to amplify the junction region between *EF1416* and *EF1490* and between *EF2083* and *EF2146*, respectively.

Data analysis. The bimodal distribution of hybridization signals highlights the existence of two different gene populations corresponding in fact to their presence or absence. Each gene population fits a Gaussian distribution model. The hybridization signal distribution, $f(x)$, is modeled by a mixture defined as the weighted sum of the two Gaussian distributions (one for each population): $f(x) = pN(x;\mu_1,\sigma^2_1) + (1 - p)N(x;\mu_2,\sigma^2_2)$, where x, the hybridization signal, is the log-transformed signal mean intensity, p is the proportion of genes in the first class, and $N(x;\mu_i,\sigma^2_i)$ is the Gaussian density of probability with mean μ_i and variance σ^2_i of population i ($i = 1$ or 2).

The parameters of the mixture model are estimated by maximum likelihood using an expectation maximization (EM) algorithm implemented in the library MCLUST (20, 21) available in the R statistical environment (http://www.r-project.org). The maximization step of the EM algorithm generates an estimation of the parameters (mean and variance) of each Gaussian component and of the mixing proportion p. The expectation step computes the probabilities that the observed signals belong to either class. The EM algorithm was applied using several initial values, and parameters maximizing the likelihood of the model were retained. Gene classification was based on the probability that the relevant probe belonged to the population of high or low hybridization signals (i.e., respectively, present or absent genes). The macroarray data were analyzed using Class2G (http://migale.jouy.inra.fr/class2g), an R plug-in based on the R package MCLUST and integrated in the BioArray Software Environment (49).

RESULTS AND DISCUSSION

Macroarray data analysis. In setting up comparative genomic hybridization for *E. faecalis*, we first developed a method to distinguish positive from negative hybridization signals using a normal mixture model with two Gaussian components. Comparative genomic hybridization of *E. faecalis* strains was performed on a set of 186 *E. faecalis* genes, which mainly encode putative membrane proteins, lipoproteins, cell wall surface proteins, or secreted proteins (see Table S1 in the supplemental material). We specifically focused on bacterial cell surface proteins that are known to participate in adaptation and/or survival in various environments. Among them were included *E. faecalis* known virulence genes (*gelE*, *cylB*, *cspI*, and *asa1*). We also took care to select genes at different genome locations. Total DNA isolated from 30 strains including V583 and MMH594 strains as positive controls was used for macroarray hybridization. Raw and analyzed hybridization data are available in the supplemental material (see Table S2). No hybridization was detected when *E. faecium* DNA was used as target (data not shown), showing that the conditions used for gene detection with the array were specific. In our macroarray, EF3238, encoding the beta subunit of RNA polymerase, shares the highest DNA sequence identity with the *E. faecium* counterpart (85%). Consequently, we estimated that the sequence identity of genes classified as present was above

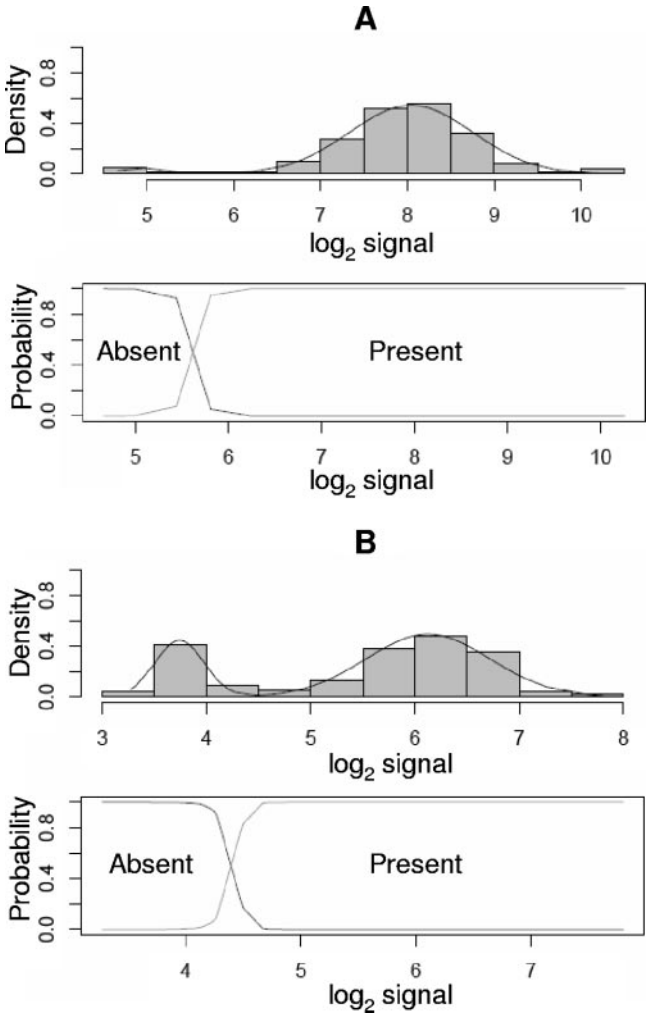

FIG. 1. Distribution and density of probability of hybridization signals modeled by a mixture model with two Gaussian components obtained with the V583 (A) and VE14531 (B) strains.

85%. According to our data and the data of other investigators (42, 44), nucleotide sequence identity between gene homologs within the *E. faecalis* species is at least 98%. The classification of genes as absent from a given isolate was therefore considered as unambiguous. Classification of a gene as present or absent often relies on empirical cutoff ratios calculated from independent experiments with a reference strain. However, due to inherent experimental variability, this analysis method leads to the misclassification of genes (36). To ensure reliable data analysis of our hybridization results—and since our primary aim was to classify genes in two groups, i.e., present and absent—we analyzed distribution of the hybridization signal of each array by a normal mixture model with two Gaussian components using Class2G software (see Materials and Methods). Hybridization signals clearly segregated into two groups, strong and weak, indicating the presence or absence of the specific target genes. As expected, hybridization signals of strain V583 (Fig. 1A) exhibited an almost normal distribution due to the fact that most of the probed genes were selected from that genome. In contrast, a bimodal distribution was observed for the tested strains, indicating that some genes were absent, as shown in Fig. 1B.

We observed some overlap between the two populations, such that an accurate threshold to classify genes as present (class 1) or absent (class 0) could not be ascertained. For final classification, genes were classified as present (probability of belonging to class 1 was ≥ 0.9), absent (probability of belonging to class 1 was ≤ 0.1) or ambiguous (probability of belonging to class 1 was between 0.1 and 0.9). This allows detection of potentially unreliable classification (ambiguous genes) that can be further investigated to ensure results. The status of the genes classified as ambiguous was established by 249 PCR amplifications. Overall, 121 genes were confirmed present in some isolates. In addition, to estimate the true-positive rate (ratio of positives correctly classified over the total positives) and the false-positive rate (ratio of negatives incorrectly classified over total negatives), we performed 1,130 specific PCR amplifications. The true-positive and the false-positive rates had an average of 84% (median, 90%) and 16% (median, 12%), respectively. False-positive misclassification may result from (i) biased choice of PCR amplifications, as half of them were performed on genes classified as ambiguous in some isolates, (ii) the attraction of the larger group (class 1), and (iii) cross-hybridizations between fragment probes and homologous genomic sequences as paralogs. Indeed, as the *E. faecalis* V583 genome carries a number of paralogs (46), it is likely that some of the genes classified as present correspond to paralogs with different cellular roles. Closer analysis of our probes revealed that probes EF0146, EF0487, EF0492, EF3076, and EF3253 may cross-hybridize as they have V583 paralogs sharing more than 89% identity at the nucleotide level. Their detection in the isolates thus indicates the presence of at least one member of the corresponding gene family.

The data analysis method used in this work is particularly adapted for comparative genomic hybridization, especially for DNA-DNA macroarray data in the context of gene classification. Its main advantage over commonly used methods is that it consists of an array-to-array analysis that limits the impact of experimental variations between strains. This is because it requires neither arbitrarily or experimentally fixed threshold values related to reference strain(s) nor extensive data normalization. Furthermore, it provides a measure of confidence in the gene classification.

Identification and distribution of the variable genes. Results of the macroarray data on 186 *E. faecalis* probes plus the 16 genes analyzed by PCR (see Material and Methods and Table S3 in the supplemental material) as tested on 30 isolates are shown in Fig. 2. Of the 202 *E. faecalis* genes, a total of 124 were detected in all 30 isolates, indicating that they are conserved. The number of absent genes in comparison to strains V583 and MMH594 varied between 68 (in food isolate VE14585) and 30 (in clinical isolate VE14514). In total, 78 genes (Table 2) proved to be variable, and their distribution in isolates was classified in three groups.

Group I. The largest group comprised 51 genes that were equally detected in both food and clinical isolates, indicating that these genes cannot be used to determine strain origins (Table 2). For example, *gelE* and *asa1*, corresponding to EF1818 and EFA0047 in V583, are characterized as virulence factors yet are detected in several food isolates. These results


VE14571 S
VE14039 C
VE14040 C
VE14599 C
VE14598 C
VE14568 C
VE14597 C
VE14596 C
VE14512 C
VE14518 C
VE14569 C
VE14583 C
VE14584 C
VE14514 C
MMH594 C
V583 C
VE14510 F
VE14615 F
VE14617 F
VE14523 F
VE14628 F
VE14522 F
VE14505I
VE14524 F
VE14623 F
VE14531 F
VE14532 F
VE14534 F
VE14535 F
VE14585 F
Genes
Genes
0002
0025
0071
0082
0093
0122
0153
0164
0235
0304
0355
0373
0376
0397
0400
0430
0443
0455
0487
0492
0511
pai0047
pai0050
pai0053
0540
0553
0577
0592
0637
0645
0662
0700
0704
0737
0753
0756
0773
0776
0799
0807
0832
0927
0965
0982
1038
1093
1099
1200
1221
1269
1335
1420
1533
1641
1683
1800
1825
1876
1896
1939
2076
2144
2168
2175
2204
2221
2234
2248
2253
2282
2373
2403
2487
2496
2513
2662
2683
2686
2728
2770
2864
2922
2970
3013
3023
3054
3074
3076
3082
3153
3155
3184
3186
3188
3198
3234
3238
3251
3253
3256
a0047
0006
0055
0077
0089
0095
0146
0163
0188
0300
0328
0361
0374
0392
0399
0429
0431
0454
0467
0490
0501
0518
pai0048
pai0052
pai0056
0552
0573
0582
0605
0638
0660
0685
0702
0713
0752
0755
0764
0775
0785
0803
0818
0886
0943
0971
1030
1090
1097
1189
1217
1231
1322
1340
1513
1596
1677
1685
1818
1844
1877
1911
1992
2080
2156
2170
2178
2216
2224
2237
2250
2254
2347
2380
2419
2488
2505
2525
2682
2684
2713
2768
2795
2903
2968
3003
3018
3041
3056
3075
3081
3106
3154
3183
3185
3187
3191
3206
3236
3248
3252
3255
3314
b0012

agree with previous studies, although a lower incidence of these genes in food isolates was reported (16, 53). Nevertheless, gelatinase activity was not detected in 38 and 50% of the food and clinical isolates of our collection, respectively (Table 1). Indeed, various isolates carrying *gelE* reportedly fail to produce gelatinase, likely due to a deletion of the *fsr* cluster region (16, 41, 48). The low frequency of gene EF1825, which is included in the 23.9-kb deletion comprising the *fsr* locus, suggests that such deletions may have occurred (Table 1). For GelE-positive strains that do not carry EF1825, the deletion may have occurred such that the *fsr* genes are still present. Similarly, genes *cpsI* (EF2487), *cpsH* (EF2488), and EF2175 belong or are close to the capsular polysaccharide biosynthesis gene clusters *cps* (EF2485 to EF2495) and *epa* (EF2180 to EF2200) that are implicated in *E. faecalis* virulence (29, 58, 59). These genes are also found in food isolates.

The low incidence of *asa1* and *cpsI* in clinical isolates observed here contrasts with previous studies (3, 10, 29), possibly reflecting differences in the origins of the isolates. However, the low *asa1* frequency in this study may be explained by a higher primer specificity since primers used in previous reports also matched with the other aggregation substance-encoding genes *prgB* (2, 9) and *asp1* (9) of plasmids pCF10 and pPD1, respectively. This observation may explain why isolates VE14568 and VE14571 were found negative for *asa1* compared to the study of Archimbaud et al. (3). We therefore suggest that the frequency of *asa1* has been overestimated in previous studies.

Conservation of the genes encoding factors important for virulence may suggest that selection for the maintenance of such traits may exist in their natural environment, which might be as diverse as soil or plants, and in insect, reptile, bird, or mammal digestive tracts (for a review, see reference 2). Although all the above factors are known to enhance virulence in animal or cellular models, they may not be sufficient for *E. faecalis* pathogenicity.

Group II. The variable genes of the second group were more often found in clinical isolates than in food isolates. Thirteen were absent in the dairy food isolates, and two genes were more frequent in clinical isolates (Table 2). Genes EF0552 and EF0553 belong to a putative operon encoding a xylose-containing oligosaccharide phosphotransferase system (PTS) transporter (60) and could confer the ability to colonize particular biotopes, such as plants, that provide xylose-rich polymers.

We examined 70 additional enterococcal isolates to confirm the classification of genes scored as absent from a total of 50 dairy food strains. Specific PCR analysis was performed using gene-specific primer pairs. Interestingly, *cylB* that codes the ABC transporter of cytolysin, turned out to be present in a particular class of food isolates corresponding to ewe and goat cheese isolates (Table 3). The higher rates of *cylB* in food isolates in previous studies may be due to sampling enriched in these food groups (16, 52, 53). Gene EF2170, which is part of the *cps* cluster, was significantly underrepresented in dairy food isolates. Differential distribution of *cps* genes could impact the cell wall polysaccharide composition and thereby contribute to serological differences among isolates (31, 32).

Six of the PCR-probed genes were not detected in any of the dairy food isolates tested, whereas they were found in 13 to 21% of the clinical or commensal isolates (Table 3). Interestingly, three of them (EF0573, EF0592, and EF0605) were codetected in the same clinical isolates, suggesting that the isolates may derive from a common ancestor. Among these, EF0592, an adhesin-like encoding gene, was reported to be exclusively associated with clinical isolates (10). Genes EF1420 and EF2144 encode two nonhomologous putative lipoproteins that are specific to *E. faecalis* and are embedded within prophages. When tested in several food isolates, the entire prophages were absent (see below), raising the possibility that up to 139 genes are coordinately absent. To our knowledge, *E. faecalis* phages have been scarcely studied; however, it is tempting to speculate that, as previously proposed for several bacteria, prophages may encode fitness factors that confer a benefit under peculiar ecological conditions (4).

This is the first report of identification of *E. faecalis* genes absent from food isolates. Their low incidence in clinical and commensal isolates suggests that these genes may act as putative fitness factors. However, their detection in food isolates may be a valuable indicator of potential risk.

Group III. The third group comprises genes absent in all isolates except reference strain V583 and related isolate MMH594 (Table 3). Eight genes (excepting V583-specific EF2513) may be good markers for the common lineage of MMH594 and V583. Some of them may be particularly mobile as they are close to or part of putative mobile elements (compiled in Table 4). Since bacterial pathogenesis is a multifactorial process, strain-specific genes may contribute to the epidemic spread of this lineage. These genes are likely to constitute clonal markers that should be traced to follow their spread among isolates.

Strain variability and gene conservation. Among the genes analyzed in this work, 61% (124/202) were shared by all isolates tested and may be part of the "core" set of *E. faecalis* genes. Conservation of putative adhesion proteins suggests that they may be related to a global requirement for adhesion to survive in different ecological niches. Among these, the cell wall-anchored proteins with tandem repeats of the immunoglobulin fold, encoded by the EF0089, EF1093, EF1099, EF1269, and EF2224 genes, were recently reported to be ubiquitous among *E. faecalis* (43, 55). The high-affinity dicarboxylate carbohydrate transport system encoded by the EF0429-EF0431 operon is rarely found in gram-positive bacteria but is totally conserved in *E. faecalis*. In the V583 genome, it is

FIG. 2. Detection of a selection of *E. faecalis* genes in 30 isolates from clinical and food origins. The individual chromosomes are displayed vertically, and genes are ordered according to their organization in the reference strain V583 and in MMH494 PAI. The names of the strains are indicated on the top. C, F, and S indicate, respectively, clinical, food, and human stool origins of the isolates. Gene designations are on both sides of the diagram, with one of every two genes identified alternately on the left and the right. Absent and ambiguous genes are indicated by white and gray squares, respectively. Bold characters indicate variable genes.

TABLE 2. Distribution of 78 genes identified as variable between *E. faecalis* isolates

Gene group	ORF	Gene name	Gene product	% GC content	Predicted location	Detection in[a]	
						Food isolates (% [n = 12])	Clinical isolates (% [n = 15])
Group I	EF0122		Conserved domain protein	30.4	Secreted	33	33
	EF0146		Surface exclusion protein	37.7	Cell envelope	58	27
	EF0153		Cell wall surface anchor family protein	43.5	Cell envelope	25	13
	EF0163		Lipoprotein	34.9	Cell envelope	17	13
	EF0164		Lipoprotein	27.7	Cell envelope	17	7[b]
	EF0304		Lipoprotein	33.9	Cell envelope	8	13[b]
	EF0328		Conserved hypothetical protein	44.1	Cytoplasm	17	27
	EF0355		Endolysin	39.3	Secreted	67	60
	EF0376		Hypothetical protein	38.4	Secreted	100	93
	EF0487		Conserved domain protein	32.0	Secreted	42	27
	EF0490		Cell wall surface anchor family protein	41.9	Cell envelope	33	40
	EF0492		Hypothetical protein	35.8	Membrane or secreted	42	40
	EF0511		Thermonuclease precursor	39.5	Secreted	42	40
	EF0540		*N*-Acetylmannosamine-6-phosphate epimerase	34.6	Secreted	50	60
	EF0577		Adhesion lipoprotein	33.0	Cell envelope	33	73[c]
	EF0582		Membrane protein	34.7	Membrane	33	73[c]
	EF0685		Rotamase family protein	36.0	Cell envelope	92	100′
	EF0775		Gram-positive anchor protein	29.3	Cell envelope	75	100′
	EF0818		Polysaccharide lyase, family 8	37.8	Secreted	92	100′
	EF0965		Conserved hypothetical protein	28.8	Membrane	25	27
	EF1677		Lipoprotein	35.4	Cell envelope	100	93
	EF1818	*gelE*	Coccolysin	37.9	Secreted	100	80
	EF1825		Conserved domain protein	37.8	Membrane	8	20
	EF1844		Hypothetical protein	27.8	Cell envelope	42	33
	EF1876		Lipoprotein, NLP/P60 family	41.2	Cell envelope	0[d]	13[b,d]
	EF1877		Membrane protein	40.0	Membrane	0[d]	13[b,d]
	EF1992		Endolysin	37.5	Secreted	17	40
	EF2175		LicD-related protein	35.9	Cytoplasm	25	27
	EF2234		Sugar ABC transporter, sugar-binding protein	36.7	Cell envelope	67	87
	EF2237		Lipoprotein	31.5	Cell envelope	58	73
	EF2487	*cpsI*	UDP-galactopyranose mutase	35.7	Cytoplasm	25	27
	EF2488	*cpsH*	Lipoprotein	29.3	Cell envelope	25	27
	EF2505		Cell wall surface anchor family protein	37.7	Cell envelope	83	87
	EF2525		Cell wall surface anchor family protein	36.4	Cell envelope	33	40
	EF2662		Choline binding protein	36.4	Cell envelope or secreted	83	73
	EF2682		Conserved hypothetical protein	34.7	Cell envelope	83	87
	EF2683		Conserved hypothetical protein	39.4	Secreted	83	87
	EF2684		Conserved hypothetical protein	38.5	Secreted	83	87
	EF2686		Internalin protein family	38.4	Secreted	83	87
	EF2713		Cell wall surface anchor family protein	33.5	Cell envelope	17	20
	EF2795		LysM domain lipoprotein	33.8	Cell envelope	67	53
	EF2922		Conserved hypothetical protein	33.0	Membrane	100	87
	EF2968		Cell wall surface anchor family protein	39.1	Cell envelope	100	93
	EF3023		Polysaccharide lyase, family 8	38.1	Cell envelope	83	80
	EF3074		Hypothetical protein	41.1	Secreted	75	87
	EF3075		Hypothetical protein	42.0	Secreted	75	73
	EF3154		Conserved hypothetical protein	41.4	Secreted	17	27
	EF3248		Hypothetical protein	43.2	Secreted	17	13[b]
	EFA0047	*asa1*	Aggregation substance Asa1	38.6	Cell envelope	42	13[b]
	EFB0012	*prgC*	Surface protein PrgC	42.3	Cell envelope	58	33
	PAIEF0052[e]		Nisin-resistance like protein	29.0	Cytoplasm	50	47
	PAIEF0053		Hypothetical protein	27.0	Membrane	50	47
	PAIEF0056	*esp*	Enterococcal surface protein	37.0	Cell envelope	50	53
Group II	EF0501		Lipoprotein	33.3	Cell envelope	0	27
	EF0518		Cell wall surface anchor family protein	33.0	Cell envelope	0	20
	EF0552		PTS system, IIC component	34.5	Membrane	17	60
	EF0553		PTS system, IID component	32.2	Membrane	17	60
	EF0573		Hypothetical protein	33.3	Secreted	0	20
	EF0592		Cell wall surface anchor family protein	44.7	Cell envelope	0	20
	EF0605		Conserved hypothetical protein	26.2	Membrane	0	20
	EF1420		Hypothetical protein	32.5	Cell envelope	0	33

Continued on following page

TABLE 2—*Continued*

Gene group	ORF	Gene name	Gene product	% GC content	Predicted location	Detection in[a]	
						Food isolates (% [n = 12])	Clinical isolates (% [n = 15])
	EF1896		Cell wall surface anchor family protein	40.0	Cell envelope	0	33
	EF2144		Lipoprotein	33.2	Cell envelope	0	20
	EF2168		LicD1 protein	31.4	Cell envelope	0	20
	EF2170		Glycosyl transferase, group 2 family protein	30.4	Cell envelope	0	20
	PAIEF0047	*cylB*	Cytolysin B transport protein, CylB	27.0	Membrane	0	13
	PAIEF0048	*cylA*	Cytolysin activator, CylA	31.0	Cell envelope	0	13
	PAIEF0050		Hypothetical protein	30.0	Cytoplasm	0	13
Group III	EF2248		Hypothetical protein	35.8	Secreted	0	13[b]
	EF2250		Conserved domain protein	32.9	Secreted	0	13[b]
	EF2253		Conserved hypothetical protein	38.0	Cell envelope	0	13[b]
	EF2254		Conserved hypothetical protein	42.6	Secreted	0	13[b]
	EF2282		Conserved domain protein	32.0	Cytoplasm	0	13[b]
	EF2347		Cell wall surface anchor family protein	37.9	Cell envelope	0	13[b]
	EF2513		Lipoprotein	33.0	Cell envelope	0	7[b]
	EF3153		Conserved hypothetical protein	46.5	Secreted	0	13[b]
	EF3155		Conserved hypothetical protein	40.5	Secreted	0	13[b]
	EF3252		Hypothetical protein	41.5	Secreted	0	13[b]

[a] Distribution of individual genes was analyzed by a Fisher exact test and $P < 0.05$ was considered significant. Only isolates from dairy foods were considered. Results of poultry food isolate VE14510, stool isolate VE14571 and type strain were not included.
[b] Gene was detected only in reference strains V583 and/or MMH594 meaning it was not detected in the type strain, the stool isolate VE14571 and the food isolate VE14510.
[c] This difference ($P = 0.0574$) was close to the significance level.
[d] Genes EF1876 and EF1877 were detected in the type strain.
[e] Genes PAIEFxxxx correspond to EFxxxx genes in PAI of strain MMH594 (53).

located in a region rich in putative carbohydrate metabolism genes that may encode important functions in particular environments. Regarding the diversity of the natural environments of *E. faecalis*, conserved genes are probably inherent to certain of its lifestyles. Some, if not all, of these genes may be required for gastrointestinal lifestyles in humans and/or animals.

Cluster analysis on the set of the 78 variable genes revealed three strain clusters (Fig. 3). Two clusters comprised isolates of different geographical and environmental origins. Based on our data, the urine isolate VE14514 from France is related to MMH594 and V583, mainly due to the conservation of the PAI genes probed, indicating that the PAI might be spread worldwide (43). Interestingly, besides sharing PAI genes, strain clusters I and III share EF2168, EF2170, and EF2175 and lack the putative operon EF2682-EF2686 and EF3023 and EF3075, respectively (Fig. 2). We also identified loci that distinguish the two isolates MMH594 and V583. Strain MMH594 was found to lack genes EF0153, EF0163, EF0164, EF2237, and EF2513. This suggests that V583 may have diverged from MMH594, not only by the acquisition of *vanB* and the deletion of the 17-kb region which encompass *esp* (43, 54) but also by acquisition of at least part of the mobile element efaC2 (Table 4). In addition, detection of EF2282 and EF2347 in MMH594 indicates that *vanB* genes have been transferred in an existing mobile element, leading to a complex structure identified as a putative integrative and conjugative element (5). This suggests that, independently of the evolution of the PAI region, other genes clearly contribute to define *E. faecalis* lineages.

TABLE 3. Distribution of 11 genes among a collection of 100 isolates[a]

Gene[b]	Isolate source (% of total no. of isolates)		
	Clinic (n = 38)	Food (n = 52)	Human stool (n = 10)
EF0501	18.4	5.8	0
EF0518	15.8	5.8	20
EF0573	**13.1**	**0**	**0**
EF0592	**13.1**	**0**	**0**
EF0605	**15.8**	**0**	**0**
EF1420	**21.6**	**0**	**20**
EF1896	31.6	21.2	10
EF2144	**15.8**	**0**	**10**
EF2170	**26.3**	**1.9**	**0**
cylB	8	15.4	10
PAIEF0050	**13.1**	**0**	**0**

[a] Distribution of individual genes was determined by PCR and analyzed by a Fisher exact test. Values in boldface indicate significant results between dairy food and clinical isolates with a P value of <0.05.
[b] Genes *cylA* and EF2168 were not probed as they are putatively cotranscribed with *cylB* and EF2170, respectively.

Not surprisingly, comparative genome hybridization reveals greater strain diversity than MLVA. Although we did distinguish food isolates by the absence of six genes, these genes were insufficient to cluster the isolates in a separate category (Fig. 3). These results suggest that *E. faecalis* isolates may be intrinsically related due to the fact that food isolates are likely to result from fecal animal contamination. Food isolates may have factors for establishment in their natural host, making it difficult to distinguish them from those of clinical origin.

Global genetic plasticity relative to the reference genome. When examined in the V583 genome, half of the variable genes are scattered over the chromosome, while 38 were adjacent or located in discrete regions predicted to be mobile genetic elements (Table 4).

Two genes (EF1420 and EF2144) tested in this work were

TABLE 4. Putative mobile genetic elements predicted in V583 genome[a]

Genes	Putative mobile genetic element	Integration site	Reference
EF0127-EF0166	efaC2	3′ end of EF0167 encoding a GMP-synthase	5
EF0303-EF0355	Phage01	3′ end of EF0322 encoding aminopeptidase C	46
EF0479-EF0628	PAI	Downstream of EF0477 encoding an unknown protein	54
EF1276-EF1293	Phage02	Not found	46
EF1417-EF1489	Phage03	3′ end of EF1416 encoding a glucose-6-phosphate isomerase[b,c]	46
EF1847-EF1895	efaB5	Not found	5
EF1988-EF2043	Phage04	5′ end of EF1987	46
EF2084-EF2145	Phage05	Downstream of 3′ end of tRNA-Thr2[d]	46
EF2277-EF2346	*vanB* vancomycin resistance region	Not found	46
EF2512-EF2546	efaC1	tRNA-Thr3	5
EF2798-EF2855	Phage06	Upstream of 5′ end of tRNA-Ser5	46
EF2936-EF2955	Phage07	Upstream of 5′ end of EF2935 encoding a protein of the xanthine/uracil permease family	46

[a] Genome accession number AE016830.
[b] Reported by Burrus et al. (5).
[c] *attL*/*attR*: ACAAACGCAACATGTTCGCTTTATTAGGTAAACCAGG(A/T)TTTGAAGAATTAGCAAAAGATTTAAATGCACGC(C/T)TATA (the present study).
[d] *attL*/*attR*: GGCAGGTGGCT(C/T)TTT (the present study).

located in putative prophages. Interestingly, they were absent in all food isolates tested (group II). Their absence may correspond to the absence of the cognate prophage. To further investigate the corresponding regions, we designed oligonucleotides to amplify a fragment spanning the predicted insertion site (46). Sequence analysis of the amplicons from strains that lacked EF1420 and EF2144 allowed us to identify two deletions of 48,373 bp and 43,518 bp, respectively. Consequently, we could delineate the integration site of prophages 03 and 05 of V583 strain. Prophage 03 was integrated between EF1416 and EF1490 in the 3′ end of EF1416, which encodes a glucose-6-phosphate isomerase. Prophage 05 was integrated in the 3′ end of a gene encoding $tRNA^{Thr2}$. Prophages 03 and 05 are flanked by a 76-bp repeat and a 15-bp repeat, respectively, corresponding to their attachment sites (Table 4). These results demonstrate that prophages 03 and 05 may have been acquired recently since they are absent in several natural strains. These two prophages, which were not detected in the analyzed dairy food isolates, account for a total of 135 genes. Most of them are of unknown function, and, as stated above, they may contribute to *E. faecalis* adaptation to various environments including the host. Further investigation of their distribution is required to ascertain their absence in food isolates.

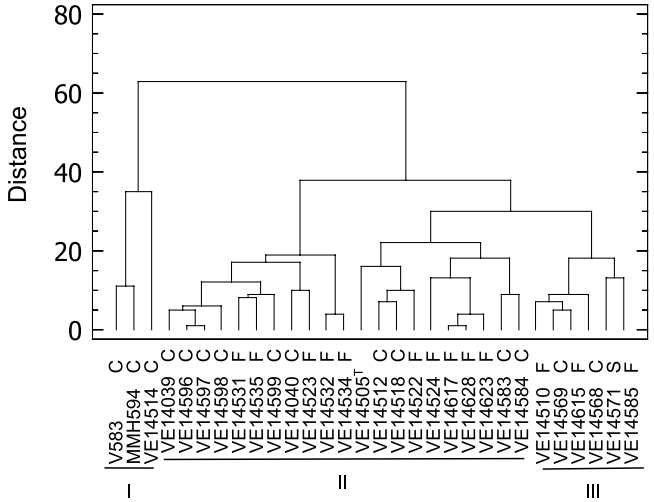

FIG. 3. Dendrogram showing the relationship between 30 *E. faecalis* isolates. Clustering analysis was performed using the StatGraphics program. The gene dissimilarity used is the squared Euclidean distance obtained from a binary matrix (genes × isolates) with 1 for the presence and 0 for the absence matrix. The hierarchical clustering was performed using the maximum linkage on the dissimilarity matrix. C, F, and S indicate, respectively, clinical, food, and human stool origins of the isolates.

The *E. faecalis* PAI region is of particular interest since it was clearly associated with two isolates responsible for multiple infections (54). Our data confirmed that the putative pathogenicity island exhibits a high variability and is largely disseminated among isolates, in keeping with a recently published study (43). In this study, 8 of the 30 strains lacked the 20 genes present in the PAI of MMH594. Several PAI genes (PAIEF0053, *esp*, EF0540, EF0577, and EF0582) were detected in about half of the isolates; others (EF0501, EF0518, *cylB*, *cylA*, PAIEF0050, EF0573, EF0592, and EF0605) had a lower incidence. This variability confirms the modular structure of this mobile genetic element. In contrast to a scenario where PAI would have evolved through deletion events only, our data suggest that PAI may evolve by gain and loss of partial PAI segments, as suggested by Shankar and coworkers (54).

All these results demonstrate that the genetic variability of *E. faecalis* isolates relates to gene loss and acquisition to not only the PAI region but also other genomic regions including capsular polysaccharide clusters, prophages, and putative mobile genetic elements. Analysis of the percent GC content of variable genes (Table 2) suggests that several of them have been recently acquired by horizontal transfer from distant organisms like gram-negative bacteria. However, the majority may result from intraspecies transfer. The efficient *E. faecalis* conjugation systems are likely to facilitate formation of variants leading to genome flexibility. As an opportunistic bacteria,

it is likely that *E. faecalis* pathogenesis results from the coordinated expression of diverse fitness and virulence factors favoring its adaptation to the hostile environment of the host and its antibacterial defenses (6). We speculate that *E. faecalis* genes identified as overrepresented in or specific to clinical isolates may constitute putative fitness factors by conferring an advantage to *E. faecalis* to establish in humans. Further genome sequencing of multiple *E. faecalis* isolates from various origins would enhance our progress in the identification of specific sequences that may encode adaptation factors to adverse environmental conditions and contribute to enhanced pathogenicity.

In summary, this is the first large-scale study of *E. faecalis* genome diversity that gives a first picture of the stable versus variable regions in the *E. faecalis* chromosome. We applied a more robust statistical method for gene distribution analysis by macroarray, which could be of general use. Several discrete regions of variability were identified, including two V583 putative prophages whose the integration sites were mapped. Our data confirm the modular structure of the PAI region with subregions conserved in more that half of the isolates. A significant finding of this work is that six of the probed genes appear to be absent from the dairy food isolates. Even if these genes are not ubiquitous in clinical isolates, they may constitute good markers for risk assessment regarding *E. faecalis* isolates that are found in fermented products. Genes specifically detected or overrepresented in clinical isolates are of particular interest as they may be fitness factors contributing to the development of human infection. Since *E. faecalis* pathogenesis is a multifactor process, we believe that extended genomic studies will allow identification of the genes needed for *E. faecalis* colonization ability and pathogenicity.

ACKNOWLEDGMENTS

We are very grateful to A. Gruss, M. A. Petit from UBLO, and F. Rodolphe from MIG for careful reading of the manuscript. We thank P. Y. Allouch, M. Gilmore, M. C. Montel, J. Anba, D. B. Clewell, J. J. Gratadoux, and E. Vandenesch for providing the isolates; we thank V. Loux and R. Bossy for genome analysis using the AGMIAL package. P.S. is indebted to A. Gross for her support.

This work was supported by the Institut National de la Recherche Agronomique and the Agence Française de Sécurité Sanitaire et Environnementale. S.B. was supported by a doctoral fellowship from the Institut National de la Recherche Agronomique and the Région Ile-de-France.

REFERENCES

1. **Aakra, A., H. Vebo, L. Snipen, H. Hirt, A. Aastveit, V. Kapur, G. Dunny, B. Murray, and I. F. Nes.** 2005. Transcriptional response of *Enterococcus faecalis* V583 to erythromycin. Antimicrob. Agents Chemother. **49:**2246–2259.
2. **Aarestrup, F. M., P. Butaye, and W. Witte.** 2002. Nonhuman reservoirs of enterococci, p. 55–99. *In* M. S. Gilmore, D. B. Clewell, P. Courvalin, G. M. Dunny, B. E. Murray, and L. B. Rice (ed.), The enterococci: pathogenesis, molecular biology, and antibiotic resistance. ASM, Washington, D.C.
3. **Archimbaud, C., N. Shankar, C. Forestier, A. Baghdayan, M. S. Gilmore, F. Charbonne, and B. Joly.** 2002. In vitro adhesive properties and virulence factors of *Enterococcus faecalis* strains. Res. Microbiol. **153:**75–80.
4. **Brussow, H., C. Canchaya, and W.-D. Hardt.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. **68:**560–602.
5. **Burrus, V., G. Pavlovic, B. Decaris, and G. Guedon.** 2002. The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. Plasmid **48:**77–97.
6. **Casadevall, A., and L. A. Pirofski.** 1999. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. Infect. Immun. **67:**3703–3713.
7. **Chan, K., S. Baker, C. C. Kim, C. S. Detweiler, G. Dougan, and S. Falkow.** 2003. Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar Typhimurium DNA microarray. J. Bacteriol. **185:**553–563.
8. **Chen, T., Y. Hosogi, K. Nishikawa, K. Abbey, R. D. Fleischmann, J. Walling, and M. J. Duncan.** 2004. Comparative whole-genome analysis of virulent and avirulent strains of *Porphyromonas gingivalis*. J. Bacteriol. **186:**5473–5479.
9. **Clewell, D. B., P. K. Tomich, M. C. Gawron-Burke, A. E. Franke, Y. Yagi, and F. Y. An.** 1982. Mapping of *Streptococcus faecalis* plasmids pAD1 and pAD2 and studies relating to transposition of Tn*917*. J. Bacteriol. **152:**1220–1230.
10. **Creti, R., M. Imperi, L. Bertuccini, F. Fabretti, G. Orefici, R. Di Rosa, and L. Baldassarri.** 2004. Survey for virulence determinants among *Enterococcus faecalis* isolated from different sources. J. Med. Microbiol. **53:**13–20.
11. **Dean, N., and A. E. Raftery.** 2005. Normal uniform mixture differential gene expression detection for cDNA microarrays. BMC Bioinformatics **6:**173.
12. **Dobrindt, U., F. Agerer, K. Michaelis, A. Janka, C. Buchrieser, M. Samuelson, C. Svanborg, G. Gottschalk, H. Karch, and J. Hacker.** 2003. Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. J. Bacteriol. **185:**1831–1840.
13. **Doumith, M., C. Cazalet, N. Simoes, L. Frangeul, C. Jacquet, F. Kunst, P. Martin, P. Cossart, P. Glaser, and C. Buchrieser.** 2004. New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics and DNA arrays. Infect. Immun. **72:**1072–1083.
14. **Dupre, I., S. Zanetti, A. M. Schito, G. Fadda, and L. A. Sechi.** 2003. Incidence of virulence determinants in clinical *Enterococcus faecium* and *Enterococcus faecalis* isolates collected in Sardinia (Italy). J. Med. Microbiol. **52:**491–498.
15. **Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos.** 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. Proc. Natl. Acad. Sci. USA **99:**1556–1561.
16. **Eaton, T. J., and M. J. Gasson.** 2001. Molecular screening of *Enterococcus* virulence determinants and potential for genetic exchange between food and medical isolates. Appl. Environ. Microbiol. **67:**1628–1635.
17. **Elsner, H. A., I. Sobottka, D. Mack, M. Claussen, R. Laufs, and R. Wirth.** 2000. Virulence factors of *Enterococcus faecalis* and *Enterococcus faecium* blood culture isolates. Eur. J. Clin. Microbiol. Infect. Dis. **19:**39–42.
18. **Facklam, R. R., M. G. S. Carvalho, and L. M. Teixeira.** 2002. History, taxonomy, biochemical characteristics, and antibiotic susceptibility testing of enterococci, p. 1–54. *In* M. S. Gilmore, D. B. Clewell, P. Courvalin, G. M. Dunny, B. E. Murray, and L. B. Rice (ed.), The enterococci: pathogenesis, molecular biology, and antibiotic resistance. ASM Press, Washington, D.C.
19. **Fouet, A., and A. L. Sonenshein.** 1990. A target for carbon source-dependent negative regulation of the *citB* promoter of *Bacillus subtilis*. J. Bacteriol. **172:**835–844.
20. **Fraley, C., and A. E. Raftery.** 1999. MCLUST: software for model-based cluster analysis. J. Classification **16:**297–306.
21. **Fraley, C., and A. E. Raftery.** 2002. Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. **97:**611–631.
22. **Franke, A. E., and D. B. Clewell.** 1981. Evidence for a chromosome-borne resistance transposon (Tn*916*) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid. J. Bacteriol. **145:**494–502.
23. **Franz, C. M., A. B. Muscholl-Silberhorn, N. M. Yousif, M. Vancanneyt, J. Swings, and W. H. Holzapfel.** 2001. Incidence of virulence factors and antibiotic resistance among enterococci isolated from food. Appl. Environ. Microbiol. **67:**4385–4389.
24. **Garrett, E. S., and G. Parmigiani.** 2003. Statistical methods for qualitative analysis of gene expression, p. 362–387. *In* G. Parmigiani, E. S. Garrett, R. A. Irizarry, and S. L. Zeger (ed.), The analysis of gene expression data. Spinger-Verlag, New York, N.Y.
25. **Gauduchon, V., L. Chalabreysse, J. Etienne, M. Celard, Y. Benito, H. Lepidi, F. Thivolet-Bejui, and F. Vandenesch.** 2003. Molecular diagnosis of infective endocarditis by PCR amplification and direct sequencing of DNA from valve tissue. J. Clin. Microbiol. **41:**763–766.
26. **Gentry-Weeks, C., M. Estay, C. Loui, and D. Baker.** 2003. Intravenous mouse infection model for studying the pathology of *Enterococcus faecalis* infections. Infect. Immun. **71:**1434–1441.
27. **Gilmore, M. S., P. S. Coburn, S. R. Nallapareddy, and B. E. Murray.** 2002. Enterococcal virulence, p. 301–354. *In* M. S. Gilmore, D. B. Clewell, P. Courvalin, G. M. Dunny, B. E. Murray, and L. B. Rice (ed.), The enterococci: pathogenesis, molecular biology, and antibiotic resistance. ASM Press, Washington, D.C.
28. **Giraffa, G.** 2003. Functionality of enterococci in dairy products. Int. J. Food Microbiol. **88:**215–222.
29. **Hancock, L. E., and M. S. Gilmore.** 2002. The capsular polysaccharide of *Enterococcus faecalis* and its relationship to other polysaccharides in the cell wall. Proc. Natl. Acad. Sci. USA **99:**1574–1579.
30. **Hancock, L. E., and M. S. Gilmore.** 2000. Pathogenicity of enterococci, p. 251–258. *In* V. A. Fischetti (ed.), Gram-positive pathogens. American Society for Microbiology, Washington, D.C.
31. **Huebner, J., Y. Wang, W. A. Krueger, L. C. Madoff, G. Martirosian, S.**

Boisot, D. A. Goldmann, D. L. Kasper, A. O. Tzianabos, and G. B. Pier. 1999. Isolation and chemical characterization of a capsular polysaccharide antigen shared by clinical isolates of *Enterococcus faecalis* and vancomycin-resistant *Enterococcus faecium*. Infect. Immun. **67:**1213–1219.

32. **Hufnagel, M., L. E. Hancock, S. Koch, C. Theilacker, M. S. Gilmore, and J. Huebner.** 2004. Serological and genetic diversity of capsular polysaccharides in *Enterococcus faecalis*. J. Clin. Microbiol. **42:**2548–2557.
33. **Huycke, M. M., and M. S. Gilmore.** 1995. Frequency of aggregation substance and cytolysin genes among enterococcal endocarditis isolates. Plasmid **34:**152–156.
34. **Huycke, M. M., C. A. Spiegel, and M. S. Gilmore.** 1991. Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*. Antimicrob. Agents Chemother. **35:**1626–1634.
35. **Kayaoglu, G., and D. Orstavik.** 2004. Virulence factors of *Enterococcus faecalis*: relationship to endodontic disease. Crit. Rev. Oral Biol. Med. **15:** 308–320.
36. **Kim, C. C., E. A. Joyce, K. Chan, and S. Falkow.** 2002. Improved analytical methods for microarray-based genome-composition analysis. Genome Biol. **3:**RESEARCH0065. [Online.] http://genomebiology.com/2002/3/11/research/0065.
37. **Koide, T., P. A. Zaini, L. M. Moreira, R. Z. Vencio, A. Y. Matsukuma, A. M. Durham, D. C. Teixeira, H. El-Dorry, P. B. Monteiro, A. C. da Silva, S. Verjovski-Almeida, A. M. da Silva, and S. L. Gomes.** 2004. DNA microarray-based genome comparison of a pathogenic and a nonpathogenic strain of *Xylella fastidiosa* delineates genes important for bacterial virulence. J. Bacteriol. **186:**5442–5449.
38. **McLachlan, G. J., R. W. Bean, and D. Peel.** 2002. A mixture model-based approach to the clustering of microarray expression data. Bioinformatics **18:**413–422.
39. **Mohamed, J. A., W. Huang, S. R. Nallapareddy, F. Teng, and B. E. Murray.** 2004. Influence of origin of isolates, especially endocarditis isolates, and various genes on biofilm formation by *Enterococcus faecalis*. Infect. Immun. **72:**3658–3663.
40. **Mundy, L. M., D. F. Sahm, and M. Gilmore.** 2000. Relationships between enterococcal virulence and antimicrobial resistance. Clin. Microbiol. Rev. **13:**513–522.
41. **Nakayama, J., R. Kariyama, and H. Kumon.** 2002. Description of a 23.9-kilobase chromosomal deletion containing a region encoding *fsr* genes which mainly determines the gelatinase-negative phenotype of clinical isolates of *Enterococcus faecalis* in urine. Appl. Environ. Microbiol. **68:**3152–3155.
42. **Nallapareddy, S. R., R. W. Duh, K. V. Singh, and B. E. Murray.** 2002. Molecular typing of selected *Enterococcus faecalis* isolates: pilot study using multilocus sequence typing and pulsed-field gel electrophoresis. J. Clin. Microbiol. **40:**868–876.
43. **Nallapareddy, S. R., H. Wenxiang, G. M. Weinstock, and B. E. Murray.** 2005. Molecular characterization of a widespread, pathogenic, and antibiotic resistance-receptive *Enterococcus faecalis* lineage and dissemination of its putative pathogenicity island. J. Bacteriol. **187:**5709–5718.
44. **Naser, S. M., F. L. Thompson, B. Hoste, D. Gevers, P. Dawyndt, M. Vancanneyt, and J. Swings.** 2005. Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. Microbiology **151:**2141–2150.
45. **Newton, M., and C. Kendziorski.** 2003. Parametric empirical Bayes methods for microarrays, p. 254–259. *In* G. Parmigiani, E. S. Garrett, R. A. Irizarry, and S. L. Zeger (ed.), The analysis of gene expression data. Spinger-Verlag, New York, N.Y.
46. **Paulsen, I. T., L. Banerjei, G. S. Myers, K. E. Nelson, R. Seshadri, T. D. Read, D. E. Fouts, J. A. Eisen, S. R. Gill, J. F. Heidelberg, H. Tettelin, R. J. Dodson, L. Umayam, L. Brinkac, M. Beanan, S. Daugherty, R. T. DeBoy, S. Durkin, J. Kolonay, R. Madupu, W. Nelson, J. Vamathevan, B. Tran, J. Upton, T. Hansen, J. Shetty, H. Khouri, T. Utterback, D. Radune, K. A. Ketchum, B. A. Dougherty, and C. M. Fraser.** 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science **299:** 2071–2074.
47. **Richmond, C. S., J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner.** 1999. Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res. **27:**3821–3835.
48. **Roberts, J. C., K. V. Singh, P. C. Okhuysen, and B. E. Murray.** 2004. Molecular epidemiology of the *fsr* locus and of gelatinase production among different subsets of *Enterococcus faecalis* isolates. J. Clin. Microbiol. **42:**2317–2320.
49. **Saal, L. H., C. Troein, J. Vallon-Christersson, S. Gruvberger, A. Borg, and C. Peterson.** 2002. BioArray software environment (BASE): a platform for comprehensive management and analysis of microarray data. Genome Biol. **3:**SOFTWARE0003.1–0003.6. [Online.] http://genomebiology.com/2002/3/8/software/0003.
50. **Sahm, D. F., J. Kissinger, M. S. Gilmore, P. R. Murray, R. Mulder, J. Solliday, and B. Clarke.** 1989. In vitro susceptibility studies of vancomycin-resistant *Enterococcus faecalis*. Antimicrob. Agents Chemother. **33:**1588–1591.
51. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. USA **97:**14668–14673.
52. **Semedo, T., M. Almeida Santos, P. Martins, M. F. Silva Lopes, J. J. Figueiredo Marques, R. Tenreiro, and M. T. Barreto Crespo.** 2003. Comparative study using type strains and clinical and food isolates to examine hemolytic activity and occurrence of the *cyl* operon in enterococci. J. Clin. Microbiol. **41:**2569–2576.
53. **Semedo, T., M. A. Santos, M. F. Lopes, J. J. Figueiredo Marques, M. T. Barreto Crespo, and R. Tenreiro.** 2003. Virulence factors in food, clinical and reference enterococci: a common trait in the genus? Syst. Appl. Microbiol. **26:**13–22.
54. **Shankar, N., A. S. Baghdayan, and M. S. Gilmore.** 2002. Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. Nature **417:**746–750.
55. **Sillanpaa, J., Y. Xu, S. R. Nallapareddy, B. E. Murray, and M. Hook.** 2004. A family of putative MSCRAMMs from *Enterococcus faecalis*. Microbiology **150:**2069–2078.
56. **Smoot, J. C., K. D. Barbian, J. J. Van Gompel, L. M. Smoot, M. S. Chaussee, G. L. Sylva, D. E. Sturdevant, S. M. Ricklefs, S. F. Porcella, L. D. Parkins, S. B. Beres, D. S. Campbell, T. M. Smith, Q. Zhang, V. Kapur, J. A. Daly, L. G. Veasy, and J. M. Musser.** 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. USA **99:**4668–4673.
57. **Titze-de-Almeida, R., R. J. L. Willems, J. Top, I. Pereira Rodrigues, R. Fonseca Ferreira, I. I., H. Boelens, M. C. C. Brandileone, R. C. Zanella, M. S. Soares Felipe, and A. van Belkum.** 2004. Multilocus variable-number tandem-repeat polymorphism among Brazilian *Enterococcus faecalis* strains. J. Clin. Microbiol. **42:**4879–4881.
58. **Xu, Y., B. E. Murray, and G. M. Weinstock.** 1998. A cluster of genes involved in polysaccharide biosynthesis from *Enterococcus faecalis* OG1RF. Infect. Immun. **66:**4313–4323.
59. **Xu, Y., K. V. Singh, X. Qin, B. E. Murray, and G. M. Weinstock.** 2000. Analysis of a gene cluster of *Enterococcus faecalis* involved in polysaccharide biosynthesis. Infect. Immun. **68:**815–823.
60. **Zuniga, M., I. Comas, R. Linaje, V. Monedero, M. J. Yebra, C. D. Esteban, J. Deutscher, G. Perez-Martinez, and F. Gonzalez-Candelas.** 2005. Horizontal gene transfer in the molecular evolution of mannose PTS transporters. Mol. Biol. Evol. **22:**1673–1685.